Journal of Elementary Science Education, Vol. 19, No. 2 (Fall 2007), pp. 25-43.

Elementary Students' Reasoning: Crests and Troughs of Learning

Anita Roychoudhury, Purdue University

Abstract

This study describes the improvement in 20 sixth grades students' reasoning abilities in the context of structured or semi-structured inquiries conducted during an after-school science club. The findings shed light on the improvement in student reasoning and on the specific areas of student difficulties. Overall reasoning skills showed more or less continuous improvement; whereas, the warrants changed in a non-linear pattern—like waves—with crests of improvement and troughs of setbacks. The study also suggests that writing played and important role in the process of student learning.

Reasoning: Crests and Troughs of Learning

There is a growing consensus of the need to educate students about scientific ways of thinking (Driver, Leach, Millar, & Scott, 1996; Hogan & Maglienti, 2001; Kuhn, 1993; Miller & Osborne, 2000; Osborne, Erduran, & Simon, 2004; Zohar & Nemet, 2002). This implies that science teaching is more than adding new information to what students already know. Teaching needs to help students learn the discursive practices of science and the scientific worldview. Teaching needs to focus on helping students understand how scientific theories are generated through the evaluation of evidence to support or refute an explanatory conclusion, model, or prediction (Suppe, 1998). Guided by this general idea, this study explores elementary students' ability to engage in evidence-based reasoning in the context of semi-structured inquiries.

Theoretical Framework

Evaluating observations and data, weighing conclusions, making informed decisions—all of these are basic thinking skills essential for becoming a responsible citizen in today's science- and technology-dependent society. Naturally, these skills have long been the focus of research in cognitive science and science education (Chinn & Brewer, 1993, 2001; Hogan & Maglienti, 2001; Kuhn, 1992, 1993; Schauble, 1990). Hogan and Maglienti (2001) and Kuhn (1992, 1993) suggest that pedagogical practices that promote coherence with prior knowledge are more likely to help students develop epistemological foundations of scientific work than simply having them engage in activities. Driver and Newton (1997) corroborate this view and recommend that science be taught as a socially constructed practice through which argumentation becomes an integral part of the discourse.

In their study of reasoning and validity testing conducted by middle school students, non-scientist adults, technicians, and scientists, Hogan and Maglienti (2001) found that those with more extensive science backgrounds exhibit different

epistemological approaches compared to the novices. Scientists strive for consistency between evidence and conclusion; coherence of conclusions with their prior knowledge is another criterion driving their decisions. Students, on the other hand, depend more on their personal views. They often make valid and invalid inferences in the context of same experiment. They do not connect their conclusions to their prior knowledge as scientists do by comparing theory and evidence. Instead, students are driven by their personal beliefs; they draw conclusions based on their ideas and incorporate the data that fits their own ideas, ignoring contradictory evidence. They treat anomalous data in an erratic manner, including it in some cases and excluding it in others, without considering the theoretical underpinnings. Other researchers have found that students can carry out the procedural aspects of an inquiry but lack the ability to carry out data interpretation, development of conclusion, and knowledge claims (Germann, Aram, & Burke, 1996; Gott & Duggan, 1995). Based on these findings, researchers propose that classrooms supporting sociocultural practices similar to those that scientists experience are likely to help develop epistemic knowledge and commitment to scientific thinking in students (Chinn & Brewer, 2001; Chinn & Malhotra, 2002; Hogan & Maglienti, 2001; Kuhn, 1993). This, in essence, highlights the need for coaching and practice in data interpretation and reasoning. Together, these studies demonstrate the importance of teaching the discursive practices that promote reasoning so that it becomes embedded in the epistemological approaches of students.

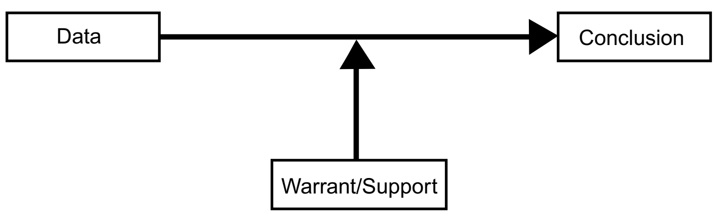
As this line of research explored the basic capacity of individuals to reason scientifically and discussed the pedagogical implications, another strand of research focused on teaching the higher level thinking underlying reasoning and argumentation (Alvermann & Hynd, 1986; Hynd & Alvermann, 1986; Osborne et al., 2004; Zohar & Nemet, 2002). Needless to say, argumentation skills require a dialogic environment in which students are required to provide support for their conclusions and choices. This was achieved via inquiry-based teaching (Zohar & Nemet, 2002), refutational texts (Alvermann & Hynd, 1986; Hynd & Alvermann, 1986), and multifarious pedagogical approaches (Osborne et al., 2004). In all cases, the instructional tasks were carefully planned to involve students in arguments and have them refute or support a position. Of these studies, Alvermann and Hynd's (1986) work with college students provides valuable insights for teaching reasoning skills through text materials; however, Osborne et al.'s and Zohar and Nemet's analyses have direct relevance to this study because they worked with students in K-12 settings and used multiple pedagogical approaches. Zohar and Nemet (2002) showed that students' reasoning skills and science knowledge improve when they are engaged in inquiries that require them to provide justification for their responses and conclusions. Explicit teaching of the principles of good argumentation resulted in considerable gain in the experimental group's performance on various measures. The researchers inferred that while students are likely to have the basic ability to develop these skills, they needed guidance and a supporting environment. The classroom culture supported the discourse of reasoning and argumentation and helped in the active construction of knowledge. They further postulated that because such an environment is generally rare in typical classrooms, students' ability to reason well does not get much of a chance to develop hence special attention needs to be given to this dimension of learning.

Although both researchers and science teachers agree that reasoning is an important aspect of the epistemology of science, it is rarely addressed in classrooms (Means & Voss, 1996; Newton, Driver, & Osborne, 1999; Scott, 1998; Zohar & Dori,

2003). The responsibility for this shortcoming does not rest solely on teachers even though they design the teaching environment this way. They learn science primarily through the transmission mode, and it is well-known that teachers teach the way they are taught (Hand & Treagust, 1994; Hewson, Tabachnick, Zeichner, & Lemberger, 1999; Lortie, 1975). Nevertheless, teacher education and its influence on pedagogy at the school level are beyond the scope of this study; here the focus is on fostering scientific reasoning in students from the early grades.

Because argumentation and scientific reasoning have been the focus of the research described earlier, it is essential that these two concepts be explained before providing a description of the study. Argumentation involves supporting or refuting one of two competing claims via weighing evidence and conclusions. This essential aspect of scientific discourse involves basic reasoning skills that help one arrive at a conclusion from a set of observations (Kuhn, 1993; Kuhn, Schauble, & Garcia-Mila, 1992; Toulmin, 1958; Toulmin, Rieke, & Janik, 1984). According to Toulmin et al. (1984), the foundation of scientific reasoning consists of claims drawn from data (ground) with the support of warrants (See Figure 1). This model of reasoning can be incorporated into science teaching pedagogy at any level even when students are not engaged in argument but simply drawing conclusions from observations. The basic model of reasoning represented in Figure 1 can help students learn science concepts by engaging them in the fundamental kernels of the discursive practice of the field.

Figure 1. Schematic Representation of Basics of Reasoning



This aspect has not been explicitly examined in the studies described above in which students were engaged in arguments about competing theories. These studies provide valuable insights, but it is essential that the basic model of reasoning from data also be explored. Furthermore, it is also necessary that teaching of such skills begin early so that students become accustomed to supporting their claims with warrants or look for them behind any claims and thus become better equipped for arguments in more complex contexts in later years of schooling. Using this assumption, this study explored elementary school students' reasoning as they explored new concepts.

Due to the major role that writing plays in learning, it was used to frame the pedagogy of reasoning adopted in this study. Researchers such as Applebee (1984), Bereiter and Scardamalia (1987), and Klein (2000) suggest that the very act of writing promotes thinking. Among these studies, Bereiter and Scardamalia's hypothesis about "writing-to-learn" has considerable influence on current educational psychology. According to this model, writing is a dialectical relationship between rhetorical and content space. In the rhetorical space, the writer defines the purpose and the audience. This, in turn, determines and shapes the content, and through writing, the writer makes the rhetorical space congruent with the content space. For

example, a student may want to set the rhetorical goal of describing the influence of temperature on fish metabolism. To attain this goal, she might set the content goal of describing the metabolic rate and the experiment she did. Another content goal would be to describe the findings and their meanings. Researchers exploring the role of writing in the context of instruction postulate that the dialectical interactions between such rhetorical and content space contribute to learning.

Other studies show the impact of various forms of writing on learning content from different angles; of course, some types of writing help more than others. Newell (1984) compared various forms of writing in a study with eight 11th grade students as they worked with science and social studies materials. The writing tools used by the students were note-taking, answering study questions, and writing an analytical essay. The only measure of outcome that showed a significant effect was the essay writing samples. Another study using fifth and sixth grade students (Laidlow, Skok, & McLaughlin, 1993) showed that note-taking helped students on weekly quizzes in science. Langer & Applebee (1987) reported on a series of three studies examining the influence of writing on learning and thinking. While the different writing tasks influenced learning in different ways, a comparison with the reading-only groups showed that any form of writing is better than simply reading the content materials. Tierney (1981) found that use of expository and expressive writing to learn biology was effective in producing superior outcomes on a delayed measure for the experimental group. Given writing's influence on learning for all of these various age groups, it was assumed that writing reports on each inquiry would further enhance the thinking and reasoning abilities of students in this study.

In addition to generating knowledge, writing also creates a discourse space in which students and teachers can interact through feedback loops. These interactions help both teachers and students modify the teaching and learning process. Written products allow teachers to get a sense of how all students are thinking—something that is difficult to achieve in everyday teaching through question-answer and discussions only. In classroom interactions, students participate unequally, so writing affords teachers the opportunity to explore every student's thinking and modify teaching accordingly.

Thus, the overall purpose of the study was to help students develop reasoning skills via structured inquiry activities and writing reports in the framework of a supporting environment. An additional goal was to explore the nature of instructional supports necessary to help students in this regard.

Method

The School and the Participants

This study was conducted in an urban school in a midwestern city as part of a larger project that involved conducting an after-school science club over the course of one academic year. Twenty girls, from a pool of sixth grade student volunteers from several different sections, were selected to participate in the science club because they had obtained written consent from their parents prior to its commencement. According to their homeroom teachers, a majority (approximately 75%) of these girls were average or below average students in their respective sections. One of the teachers speculated that these girls might have joined the club hoping to get some extra practice in science. The higher performing students, on the other hand, might be satisfied with their academic achievement

and were therefore more interested spending their time in other extra-curricular activities.

The After-School Science Club

The science club activities consisted of structured or semi-structured inquiries led by a researcher (the author) and a science education graduate student. These inquiries were designed in consultation with the teachers so that they would supplement student learning from classroom instructional activities. While the inquiries and activities were not exactly part of the school curriculum, they were related to it. The curricular pressures felt by these urban teachers due to state-mandated tests influenced this focus, and thus, the overall goal was to enhance the students' formal science knowledge by supplementing the classroom curriculum.

Another decision also influenced by the school context was the choice to use structured inquiries as an appropriate framework for the club activities; these students were unaccustomed to doing open-ended inquiries. The findings from research on inquiries performed by students of this age group also informed the decision. Keys (1998) found that sixth grade students have difficulties in designing open-ended inquiries; thus it seemed that an appropriate, effective approach would consist of some structure for the activities the students would pursue in the club. At the same time, they were allowed to try additional inquiries to satisfy their curiosity stemming from the planned ones. The focus questions for the structured inquiries were given to the students, and, in some cases, they finalized the design before carrying out their projects; in others, they had some basic guidelines. For the most part, they worked in groups of two or three, but during almost every session, there were also some large group discussions aimed at helping the students relate their learning to their personal experiences. Students often cited examples from events they had observed at home or at school and asked questions regarding their observations; on some days, these discussions constituted the better part of the session.

It was apparent from the inception of the science club that student writing was mostly sketchy and incomplete. They needed continuous encouragement and support to elaborate on their initial responses. It was also noted that while the students were learning the content, they were unable to reason clearly and often had difficulty articulating the rationale for their conclusions. This led the researchers to focus on reasoning during the later part of the year (winter and spring), and it is this particular process that is reported here.

The inquiries during this study were designed around the students' questions raised earlier in the year about acid rain. They were curious about the effects of this phenomena, and it seemed that they would benefit from exploring its effects on plants and rocks first-hand. The inquiries on the topic began with the exploration of the basic properties of acids and bases, as these students knew only the term *acid* and did not know what an acid is or how it is different from a base or a neutral substance. The primary inquiries are described in Table 1. In addition, students performed other inquiries to find answers to their own questions. For example, after testing the pH of distilled water, they decided to bring water from various places, such as the nearby river, pond, home, and other sources to test and compare their pH levels.

Table 1. Summary Description of Activities

Title	Description
Spreading Colors (Warm-Up Activity) – The aim was to explicitly teach the idea of observation-conclusion-support (warrant).	Two spots—one with blue and one with yellow fabric paint—were painted on a t-shirt. A similar set of spots was painted on another part of the t-shirt. One set of spots was sprayed with water, and then comparisons were made between the two sets for drawing conclusion.
pH Level – What are differences in the chemical properties of some common household substances?	The pH levels of the substances were tested with pH papers, and they were classified into acids, bases, and neutral substances. A small amount of acids were tasted, and bases were felt by rubbing a small amount between two fingers. Observations were followed by conclusions about the properties of acids, bases, and neutral substances.
Magical Liquid – How do acid and base react?	Color change of cabbage juice in water, dilute ammonia, and vinegar were tested to determine the property of each liquid. Vinegar, ammonia, and cabbage juice were mixed in measured amounts. Each testing in this inquiry was preceded by prediction and followed by conclusion.
Soil Acidity – Does soil have any effect on the acidity of a liquid?	Samples of various soils were tested by filtering a diluted acid through them. The pH of the acid was tested before and after each test.
Acid on Rocks – How does acid rain act on rocks?	Effects of diluted acids were tested on limestone and chalk.
Acid Rain	Each group had two potted coleus plants. To simulate rain, they decided to spray one with a diluted acid and the other with distilled water. The dilution of the acid, the number or squirts per plant, and the frequency of watering were discussed and agreed upon by all the groups. All the plants were placed in the same area to control the ambient factors. The number and general conditions of the leaves were recorded as the baseline data for the "condition" of each plant. This experiment was designed after the students learned about acids, bases, and neutral substances from their inquiry on pH level. Weekly observations were recorded in a journal.

The inquiries (see Table 1) were designed to help students develop concepts in a coherent manner. This means that the observations and conclusions from one inquiry were necessary to make sense of subsequent ones. For example, from the pH Level activity, students would learn, among other things, that the color change of the pH paper indicates the chemical nature of a substance. They also learned that acid and base combine to make a neutral substance. Understanding these concepts was necessary to make predictions and justify their conclusions in the next activity, Magic Liquid. In some cases, their experiments directly corresponded to similar situations in nature, such as the ones with plants or soil; in other cases, analogies were used to generate discussions about environmental issues.

This study was conducted over the 12-week period that the students took to complete the inquiries on acid rain. Because of the emphasis on discussions and students' own experimentation beyond what was suggested, inquiries often took longer than our projected timeframe. The inquiries culminated with the students viewing of the videotape *Acid Rain: The Invisible Threat* (Scott Resources) to see examples from nature and relate their inquiries to real situations.

Students' efforts in reasoning needed scaffolding to various degrees; initially, extensive help was provided in the form of suggestions, guiding questions, and discussions, but this help was slowly retracted over the course of time. Students needed to provide written responses to questions and prompts requiring them to draw conclusions from observations supported by warrant. Toulmin et al.'s (1984) model of reasoning from data (see Table 2) was used to guide the pedagogy of reasoning in this study and also for analyzing student reports.

Students were taught to make claims from their data and explicitly provide support for their claims. Considerable emphasis was given to this latter aspect, as it appeared to be especially challenging for the students. Earlier in the club, students explored various properties of magnets, and one of the focus questions was "Which magnet is the strongest?" The data and conclusions from one group are presented in Table 2. Evidently, the conclusion follows directly from the data, and as a result, it did not pose any difficulty for most groups. The warrant, however, was absent from all of the students' work, and such omissions were typical of almost all the students before this was specifically taught and required in their reports. It needs to be noted here that the term *warrant* was difficult for sixth grade students to understand; instead they were asked to provide *support* and *justification* for their conclusions.

Table 2. Example of Toulmin et al.'s Schema of Reasoning from an Inquiry

Data Distance from Where a Magnet Can Pull (cm)	Claims	Warrant*
Magnet A – 8.0, 7.8, 8.4 Magnet B – 5.2, 5.0, 4.6 Magnet C – 2.4, 2.6, 2.4	Magnet A is the strongest.	Magnet attracted from the longest distance in all trials

***In general, this aspect was absent from most students' reports.**

Data Collection

As indicated above, writing reports constituted a major facet of the club activities, and these reports were the primary source of data for the study. Student reports from each inquiry were collected regularly at the end of the club meetings and examined for science content understanding and the quality of reasoning. A secondary data source was the author's research journal. The author kept a journal that included observation notes from each session and reflection from a researcher's as well as an instructor's point of view. In addition, this journal also contained the author's summary of all the discussions with the graduate student who assisted in the club activities and with the teachers of the participating students. The author and the graduate student had a discussion before and after each club meeting; frequently, the teachers met with the author and provided their feedback relative to various facets of

the club. The specific processes of the analyses of the data from these two sources—student reports and author's journal—are described in the subsequent section.

Data Analysis

Students' reports were examined for their reasoning in the following manner. The number of correct conclusions in each report was marked, and then the warrants for each conclusion were examined for validity. After marking the responses, the percentage of valid conclusions and warrants for each activity out of the total possible conclusions and warrants was calculated. The percentage of valid responses across the activities over time are reported and discussed in the findings section. A sample of acceptable conclusions and warrants for each inquiry is presented in Table 3.

Table 3. Example of Acceptable Conclusions and Rationales in the Context of Each Experiment

Experiment Focus	Example of Conclusion	Example of Rationale (warrant)
Observation of differences between putting fabric color on wet (marked A) and dry (marked B) areas on fabric (cotton)	Water caused the set A colors to spread. (Roxie)	Because set B did not react the way set A did. (Roxie)
pH level to separate acids and bases followed by tasting of acids and feeling of the basic substances	Taste of acid: Most acids taste sour. (Tina)	Over half of the acids we had are sour. (Tina)
Acid-base neutralization using cabbage juice, vinegar, and ammonia	A base plus acid equals neutral. (Roxie)	It looks clear. We saw water was clear. (Valerie)
Samples of garden soil acidity testing	There is base in the soil. (Valerie)	When you pour acid in it, the liquid became less acid. (Valerie)
Acid on rocks	Acids damage rocks. (Amber)	Because the rocks changed when we put acid. (Amber)
Effects of acid rain and normal water on plant health (long-term project)	Neutral is good for the plants. (Tiffany)	Because neutral ones did pretty well and the others look pretty ugly. (Tiffany)

In addition to examining the correct reasoning, student errors and omissions, particularly their dealings of anomalous data, were also analyzed using a pattern-coding scheme (Miles & Huberman, 1994). For this part of the analysis, each report was read at least twice, with the first round of reading aimed at determining whether there was a pattern in their errors. In the second reading, each pattern was further examined for the details in the nature of errors. The next step in the analysis consisted of examining the author's journal for emerging patterns

through iterative readings. The emergent patterns were checked for elaboration, explanation, or disconfirmation of the trends noted in the student reports. The author's observation notes, as well as the summaries of the discussions among the author, the graduate student, and the teachers, were used to triangulate the findings from the students' reports.

The entire analysis was conducted by two readers: (1) the author and (2) an independent researcher, experienced in coding this kind of data but unconnected to this study. These readers met regularly to discuss the analytical strategies, compare the outcomes, and review the coding process until a consensus was reached about the cases in which the initial interpretations differed. Overall, the coding was consistent in 89% of cases, and this constituted the basis of the claims made in the Findings section.

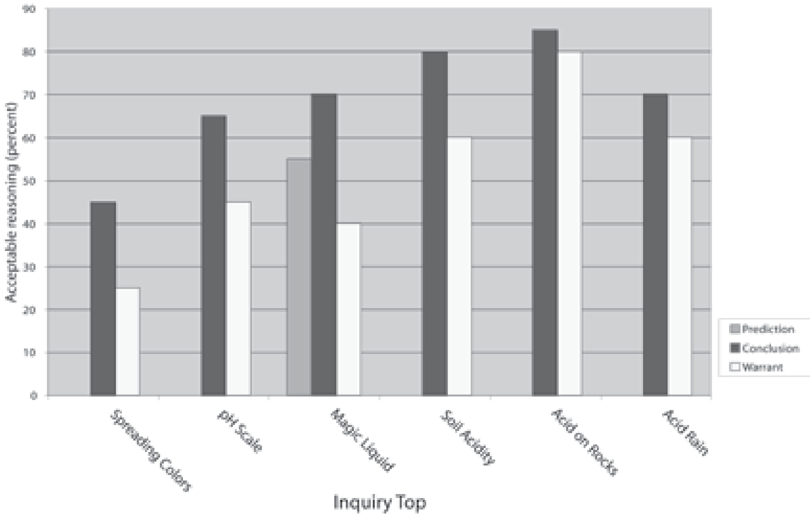
Findings

The findings from the study are organized around the three themes that emerged from the various datasets: (1) overall changes in reasoning, (2) where and what kind of guidance?, and (3) trends in errors. These themes describe the progression in student reasoning, the nature and extent of scaffolding needed, and the typical errors and their contexts within the inquiries.

Overall Changes in Reasoning

There were multiple observations and conclusions to be made in most of the inquiries, so instead of comparing the raw frequencies of acceptable responses, the percentage of valid conclusions and warrants for each activity was used to represent the overall changes in student reasoning over time. The data showed that for most of the inquiries, the percentages of warrants were less than that for the conclusions, indicating students either did not provide warrants or provided incorrect warrants for their conclusions.

Figure 2. Percentages of Scientifically Acceptable Responses per Activity



It is apparent from Figure 2 that the validity of the conclusions drawn by the students progressed in an almost linear pattern with an increase in the valid interpretations of the data in most cases with the exception of the project on acid rain. Warrants, on the other hand, progressed in a nonlinear fashion, sometimes showing an increase in total percentage and other times a decrease. A plausible explanation for this pattern lies in the nature of the activities and the extent of scaffolding provided by the researchers. Some inquiries were more complex than others, yet the researchers had begun retracting direct help at that point of the study, resulting in less improvement or even setbacks, particularly, in the warrants developed by students. It is obvious from Figure 2 that less than half of the students (45%) provided valid conclusions, and an even smaller number of students (25%) provided valid warrants for the Spreading Colors activity. This was not surprising because separating the observation and conclusion was problematic for them from the beginning, and even those who could conclude appropriately did not provide any warrant. This could be a function of the nature of the data in this particular inquiry because in previous cases of simple quantitative data involving a direct relationship between two variables, students were able to make valid conclusions. For example, during the earlier days of the science club, before this explicit focus on reasoning was adopted, students carried out an experiment with magnets, and they were able to make valid conclusions from the data (see Table 2), albeit without any warrants.

In case of qualitative observations such as the ones in the experiment with Spreading Colors, however, they had difficulty in separating the data, conclusion, and warrant. Despite the simplicity of the reasoning underlying this activity, students had difficulty in distinguishing among the three fundamental aspects of scientific reasoning. They wrote conclusions as observations and were also unable to separate warrants from conclusions. For example, Libby wrote, "water spread out the color" as her observation, and for her conclusion, she wrote that "because the colors spread out in the wet part." This pattern of responses was common

at the beginning of the activity and provided the opportunity for the researchers to explicate the difference between the observations and conclusions as well as engage students in a simple observe-conclude-justify cycle (see Figure 1). Through extensive coaching, discussions, examples, and practice with a kernel of the basic discursive practice in scientific reasoning during the session, students were able to modify their reasoning. For example, Libby's written responses that she submitted after revision had the following as her observation, conclusion, and warrant. She observed, "the color spots became bigger after we sprayed water." This was followed by her conclusion and warrant, "water made the spot spread because the dry spots did not grow." Her response is an example of the large majority (85%) who gave valid reasons, and most of these students also provided valid warrants for their conclusions after instruction.

The rate of success was different for the pH Level activity, which involved multiple observations and conclusions. During this activity, the researchers provided less direct guidance about the conclusions and warrants; consequently fewer students reasoned well. Likewise, the Magic Liquid activity was also complex in nature because it required students to make predictions based on their prior knowledge about acidic, basic, and neutral liquids from the pH scale activity. This was not easy for them, and even with some indirect guidance, only 55% students made a valid prediction. To get meaningful results, they had to be precise in performing the neutralization, and naturally, some students were not sufficiently careful about the amount of acid and base they mixed. They needed guidance and had to repeat the process a few times to obtain neutralization, and although their conclusions improved (70%) after this, the warrants were either invalid or absent. In general, providing justification for the conclusions was difficult in this inquiry. At the end, when they had to justify the conclusion that the final product was neutral, over half of the students (55%) could not provide a proper rationale; as a result, the overall reasoning from this inquiry was inferior to that from the previous experiment.

Over the next several weeks, the discussions, feedback, and additional practice in reasoning resulted in improved responses for the Soil Acidity and Acid on Rocks inquiries. A majority of these students correctly concluded (Soil Acidity: 75%; Acid on Rocks: 80%). Many of them (Soil Acidity: 60%; Acid on Rocks: 75%) also supported their conclusions with appropriate warrants, but again their reasoning deteriorated as they were bringing closure to a long-term project. They were making observations of the effects of acid rain on plants for several weeks in parallel with other short-term inquiries. At the culmination of the project, each group had to make a final observation on their own set of plants, aggregate the class data, and draw a final conclusion. This, however, proved to be a major challenge, as there was an anomaly in the data. Some of the plants receiving acid rain were doing well; whereas, the general trend indicated that the acid rain damaged the plants. The groups that had anomalous data ignored the class data and concluded from their own observations that "acid rain did not do anything to the plants" (Laurie's report). Although 70% of the students concluded appropriately, only 60% provided valid rationale. These challenges with warrants illustrated that the integration of observations from all of the data was problematic. Also, it needs to be kept in mind that by this time, they were receiving only brief and indirect suggestions from the researchers, so their reports were written more independently than before, further demonstrating that both integration of concepts and interpretation of anomalous data requires more focused teaching. This aspect is discussed in detail under "Where and What Kind of Guidance?".

In summary, the fluctuations notwithstanding, there was almost continuous improvement in student reasoning over time; their conclusions improved for most inquiries, but their warrants fluctuated without any discernable pattern. Other researchers have observed such fluctuations in the context of data interpretation as well (Kuhn et al., 1992; Siegler, 1996). Nonetheless, an alternative explanation needs to be taken into account. It is possible that student involvement in the experiments varied from week-to-week, as these were after-school sessions and other distractions were sometimes present such as music practice, a soccer tournament, approaching holidays, and so forth. On some days, the students were more distracted and engaged in more than the usual social talk about the day's extracurricular events. Consequently, on those days, they were less focused on putting effort behind their written responses.

Where and What Kind of Guidance?

From the outset, it was evident that students needed substantial guidance with regard to drawing inferences supported by warrants. During the initial stage of this study, even for a simple activity involving only a few observations leading to a direct conclusion, students had considerable difficulty explicating the components of their reasoning. These difficulties occurred along two epistemic dimensions: (1) separation of the elements of reasoning and (2) integration of knowledge. In separating the elements, the distinction between the basic elements such as observation and conclusion posed a challenge in many situations. An excerpt from the author's journal from the first week illustrates the general pattern of student interactions and their responses:

> Almost all of them had difficulty writing the observation and conclusion separately even though we had asked them to write their observations and then the conclusions. They either wrote a conclusion and observation blended together or just a conclusion. We have seen before that they are used to writing only a phrase or a word in answer to most of the questions, so it was only natural that they would have some difficulty in articulating their reasoning. Interestingly, as they were trying to write the observations and conclusions with warrants, they were also asking why we asked them to repeat the same thing. This showed a lack of understanding of the difference between the two. We explained the differences between the components of reasoning, gave some examples, and then asked them to rewrite their observations and conclusions separately so that they would become accustomed to this practice. (Author's journal: Week 1)

Students' writing improved considerably in terms of completeness. In addition, as they attempted to write complete sentences for observation, conclusion, and warrants, they needed to articulate their thoughts, and this process appeared to help them become aware of their reasoning as well as their difficulties.

Students needed guidance along the epistemological dimension of knowledge integration. In several inquiries, students needed to combine prior knowledge with new observations to draw valid inferences. For example, the inquiries on pH Level or Magic Liquid (neutralization) involved several parts, and the students had to consider multiple observations for making predictions, drawing conclusions, and generating warrants. In the Magic Liquid experiment, they had to make predictions based on their prior knowledge and then draw inferences based on

both observations from this experiment and conclusions from the previous one. From their knowledge of acids and bases, they needed to predict the outcome of mixing cabbage juice with ammonia and with vinegar. While some students (55%) were able to predict a change in the color, most of them did not provide any rationale for their predictions. Those who provided a rationale based it on their experiences in art class or use of food coloring. For example, Amy predicted that when cabbage juice is mixed with vinegar or ammonia, "It will change color. It will be like mixing colors in art class." Sara predicted that they would change colors "because it will be like putting food coloring in water." As we interacted with students during this experiment, it became apparent to us that they needed help in integrating this experiment with their knowledge of acids and bases from their previous activities. The author's journal from this week illustrates this point:

> Interestingly, the majority of the students did not make any prediction even though they were specifically instructed to make one for all the trials with acid, base, and water and provide rationales. They tend to take a piecemeal approach toward the inquiries, treating each one as a separate activity and rarely connecting them. A few students saw the connection between this and the previous inquiry, but most of them needed a lot of suggestions. As we talked to them, we found that with some help, they could recall what they learned before. After we made suggestions such as, "Think about the changes you saw with pH papers when you used them with various substances" or asked questions such as, "Do you remember the changes you saw when you used pH paper to check acids, and other substances?" there was some improvement in their predictions and the rationales. (Author's journal: Week 4)

While students' rationale for their predictions about mixing cabbage juice with ammonia and with vinegar improved significantly, their conclusions and rationale about neutralization remained weak. They predicted that acid and base would neutralize but did not provide any rationale. A majority of the students (70%) concluded that they got a neutral product, but less than half of the students were able to provide a rationale for this conclusion. These students supported their conclusions by relating them to what they learned from the activity on pH level. For example, Valerie wrote, "This liquid turned clear, which is neutral. It is like water, which is neutral." On the other hand, the indirect suggestions about drawing conclusions were inadequate for many; they needed more explicit teaching and scaffolding to help them connect their observation to their prior knowledge. Similar problems were noted in these two epistemic dimensions in the context of some other activities, pointing to the need for carefully designed instructional strategies to address student needs in these areas.

Trends in Errors and Omissions

Another difficult area for students involved interpretation of anomalous data. In some cases, students made claims by excluding anomalous data. For example, in the inquiry on pH levels, there was an acid that tasted sweet while the others were sour; likewise, there was a base that was not slippery, and, on the other hand, a slippery liquid (hand soap) was not a base. Students, however, ignored these discrepancies and concluded that "Acids are sour" and "bases feel slippery" (Gina's report: pH Level). One could argue that the conclusions are valid, and therefore, this oversight by students is immaterial. Such an argument, however,

would overlook the implications of these conclusions for their future observations of other substances. For example, this exclusion of data could lead them to infer that a substance is not acidic if it is not sour; likewise, if something does not feel slippery, they might infer that it is not basic.

A common trend in the claims involved inclusion of anomalous data resulting from some kind of error. For the Acid Rain project, some groups found that their plants receiving acid rain were doing quite well or those receiving water were not doing well—contrary to the trend in the class data—yet they based their claims on their own data ignoring the general trend. The following excerpt from the author's journal provides details about the data-interpretation phase of this project:

> Without telling them whether or not they were right or wrong, we asked them if it would be reasonable to make conclusions based on only one set of plants or if they should look at the data from the entire class. It was apparent that they know the importance of multiple sets of data for making a study reliable. They do not know the term *reliable*, but they have the general idea from their previous work, and they were comfortable with the idea of considering class data for drawing conclusions.
>
> In order to facilitate the aggregation of data, we asked them to line up all the plants receiving acid rain in an array and the ones receiving water spray in another. This helped them see that the majority of the plants on the acid rain side had visible signs of damage. Many leaves were brown and dead. For getting a crude comparison, they used a measure devised during the planning stage of this long-term project. They counted the number of undamaged leaves left and compared that with their initial count for each group of plants. We also asked the groups that had anomalous data to reflect on the discrepant nature of their data and figure out what might have happened. They speculated that they made some procedural errors. Amber and Kelly wrote that some days they might have mixed up which one should receive water spray and which one acidic spray. Lilly speculated that they might have sprayed with too much water and thereby damaged the plant. However, these students still did not integrate this reflection with the entire dataset. In the final report, most members of the four groups that had discrepant results still concluded from their own data. (Author's Journal, week 11)

In retrospect, the student claims seemed plausible because novices often have difficulty with anomaly. Researchers have found that both children, as well as adults, tend to interpret discrepant data erratically (Chinn & Brewer, 2001; Hogan & Maglienti, 2001; Kuhn et al., 1992). People include or exclude different aspects of data in order to fit them to their personal theories. Given the ubiquity of this difficulty, it became obvious that judgments about inclusion or exclusion of anomalous data require more intensive scaffolding from instructors than the indirect and limited guidance that we provided at this point in the study. The extent of student difficulty was not apparent to us in many cases until after we received the reports; consequently, our help was in the form of feedback, something that the students did not always integrate with their subsequent work.

Limitations of the Study

Before discussing the findings and their implications, it is necessary to discuss the limitations of this study so that the reader can make an informed decision about applying any aspect of it in other contexts. This study was conducted in an after-school setting but needs to be replicated in regular classroom settings to determine its effectiveness in a typical academic environment. An additional limitation arises from the single-gender sample, which shows that the girls' reasoning abilities improved during this study. Nevertheless, it seems quite reasonable to assume that a similar approach would also help boys learn to reason since existing research does not indicate much cognitive difference between males and females (Kuhn, 1991; Linn & Hyde, 1989). Another limitation of the study stems from the voluntary nature of the students' participation. These students chose to join the science club; whereas, in a classroom context, there could be an entire spectrum of students ranging from those who are apathetic to those who are keen on science, and the results could provide very different insights into student reasoning. It needs to be kept in mind, however, that the majority of the participants in the science club were average or below average according to their academic performance, meaning that the findings could be useful in designing instructions for all students in regular classrooms.

Discussion

This study sheds light on the improvement in student reasoning and on the specific areas in which they tend to have difficulties. The findings indicate that students' reasoning abilities can develop when they engage in inquiries and are required to reason explicitly and write evidence-based claims supported by warrants. A majority of the students were able to differentiate observations from conclusions and make progressively more valid conclusions. This distinction is noteworthy, as other researchers have noted similar difficulties faced by individuals in the course of data interpretation. In studies conducted with children, as well as adults, researchers (Chinn & Brewer, 2001; Kuhn, 1992, 1993) found that participants had difficulty separating data from conclusion.

It was apparent that initially even the very basic aspect of distinguishing observations from claims needed focused coaching; this implies that students are not used to this kind of analytical thinking, which in turn suggests that the fundamentals of scientific discourse are probably missing from classroom instruction. It could be argued that when the claims are fairly straightforward, students could be aware of the warrants and leave them implicit in their reasoning. It does not, however, explain the cases in which they were asked to support their claims and predictions yet omitted them. These omissions primarily stemmed from their difficulties in explicating various elements of reasoning. Student difficulties in this area underscore the need for teaching reasoning from data because if they are unable to provide a rationale for simple conclusions, then they are likely to find it difficult to compare warrants from competing claims in the case of complex arguments as observed by Osborne et al. (2004).

While overall reasoning showed more or less continuous improvement, the warrants changed in a nonlinear pattern—like waves—with crests of improvement and troughs of setbacks. This is not unexpected since the inquiries varied in complexity. Perhaps if all the activities were similar in nature, only increasing in complexity, then one could expect students' learning to become cumulative and

their reasoning to get consistently better with each new activity. The inquiries in this study, however, were designed to vary in nature in order to keep the young students' interest levels high and keep them focused on the given topic over the course of time. Furthermore, in a classroom setting, lessons and tasks often vary in nature and in cognitive demand, and this dimension enhances the applicability of the findings of this study in typical academic settings. Given the varied nature of the inquiries, some fluctuations in student reasoning ability seem natural, but these fluctuations also highlight the areas that are relatively more challenging and in which teachers' attention is needed.

Students' inability to provide valid warrants indicated a need for continuous support for a longer period of time than was provided in the club. The extensive coaching and initial guidance given to the students were slowly retracted to explore how their reasoning abilities were developing, and this was somewhat responsible for the nonuniform changes observed. Support decreased on a relatively linear basis, but the intellectual difficulty of the experiments varied over the same period. Additionally, the fluctuating changes in students' justifications indicate that the breadth and depth of scaffolding needs to be consistent with the complexity of the reasoning underlying an inquiry. This finding corroborates what others have shown (Kuhn et al., 1992; Siegler, 1996) in different contexts. Just as in this study, other researchers have also found that young people's reasoning tends to be inconsistent—valid rationales at times and being erratic in nature at other times. The finer dimensions of reasoning are complex, and their development requires more focused instruction than we provided, which implies that in planning instruction, teachers need to anticipate the level of help based on the complexity of the inquiry and the previous experience of their students. Also, prolonged practice in the epistemic elements of reasoning is necessary to generate sustained improvement in student reasoning.

Another dimension of reasoning—integration of prior knowledge with new data to construct warrants—appeared to be challenging for the students. Clearly, reasoning in the context of a set of cohesive inquiries is not enough; teachers and researchers can design a set of related inquiries and lessons, but that is not adequate for students to easily see the underlying connections among the concepts. Others (Newell, 1984; Tierney, 1981) have noted similar trends in studies with high school students. These researchers found that even older students tend to have a superficial approach to learning science concepts and attend to only discrete pieces of information in their writing instead of integrating them into a cohesive body of knowledge even though such integration is fundamental to the epistemology of science. Scientists examine new facts and data in light of their existing knowledge and attempt to integrate new learning with the extant knowledge of the field (Hogan & Maglienti, 2001), but students tend to overlook the importance of coherence in their learning. This fundamental chasm between the scientists' and students' ways of processing data shows that integration is not a common practice in school science education and deserves focused attention from teachers and educators. Needless to say, scientists' content knowledge influences their data interpretation and integration of ideas, but our study shows that, at a very simple level, students can learn to incorporate elements of epistemic practices into their inquiries as they learn to reason. The findings also demonstrate that reasoning takes considerable time to develop, and the discursive practice of science needs to be an integral part of classroom culture; it cannot be attained in a short period of time and then abandoned.

The data from this study also shows the importance of writing in learning to reason. According to Bereiter and Scardamalia's (1987) postulates about writing and learning, these students were working in a rhetorical space consisting of a goal and content in their reports. Their goal of reasoned claims was attained via the content space, which in their case consisted of articulation of the elements of reasoning based on data. In writing the reports, the students needed to think about their conclusions as well as their justifications from the data, and the repeated practice in these epistemic elements of scientific discourse, coupled with the feedback they received, appeared to have contributed to their overall reasoning abilities. Their writing also provided the researchers with space for giving constructive feedback. There was no attempt to isolate the direct influence of writing on learning, however, so no direct support for this claim could be provided here, only that writing played a role in developing student reasoning.

In addition to the implications for lessons and activities in science classrooms, the above discussion has implications for elementary teacher education as well. It is apparent that if elementary science teaching is to change, then the teachers of elementary grades need to be taught in the same way because they need to be conversant with the epistemic practice that they need to incorporate in their teaching.

Notes

This material is based upon work supported by the National Science Foundation (NSF HRD 9908776). Any opinions, findings, conclusions, or recommendations expressed in this study are those of the author and do not necessarily reflect the views of the National Science Foundation.

Acknowledgments

The author thanks Faye Flavin for her help in the data analysis.

References

Earth Science Video Library. (1992). *Acid rain: The invisible threat* (Video: Product No. 8490). (Available from Scott Resources, Inc, P.O. Box 2121, Fort Collins, CO 80522.)

Alvermann, D. E., & Hynd, C. R. (1986). Effects of prior knowledge activation modes and text structure on nonscience majors' comprehension of physics. *Journal of Educational Research, 83*, 97-102.

Applebee, A. N. (1984). Writing and reasoning. *Review of Educational Research, 54*, 577-596.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition.* Hillsdale, NJ: Lawrence Erlbaum.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science education. *Review of Educational Research, 63*, 1-49.

Chinn, C. A., & Brewer, W. F. (2001). Models of data: A theory of how people evaluate data. *Cognition and Instruction, 19*(3), 323-392.

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education, 86*(2), 175-218.

Driver, R., Leach, J., Millar, R., & Scott, P. (1996). *Young people's images of science.* Buckingham, UK: Open University Press.

Driver, R., & Newton, P. (1997, September). *Establishing the norms of scientific argumentation in classrooms.* Paper presented at the ESERA Conference, Rome.

Germann, P. J., Aram, R., & Burke, G. (1996). Identifying patterns and relationships among the responses of seventh-grade students to science process skill of designing experiments. *Journal of Research in Science Teaching, 33,* 79-99.

Gott, R., & Duggan, S. (1995). *Investigative work in the science curriculum*. Bristol, PA: Open University Press.

Hand, B., & Treagust, D. F. (1994). Teachers' thoughts about changing to constructivist teaching/learning approaches within junior secondary science classrooms. *Journal of Education for Teaching, 20*(1), 97-112.

Hewson, P. W., Tabachnick, B. R., Zeichner, K. M., & Lemberger, J. (1999). Educating prospective teachers of biology: Findings, limitations, and recommendations. *Science Education, 85*(3), 373-384.

Hogan, K., & Maglienti, M. (2001). Comparing the epistemological underpinning of students' and scientists' reasoning about conclusions. *Journal of Research in Science Teaching, 38*(6), 663-687.

Hynd, C., & Alvermann, D. E. (1986). The role of refutation text in overcoming difficulty with science concepts. *Journal of Reading, 29,* 440-446.

Keys, C. W. (1998). A study of grade six students generating questions and plans for open-ended science investigations. *Research in Science Education, 28*(3), 301-316.

Klein, P. D. (2000). Elementary students' strategies for writing-to-learn in science. *Cognition and Instruction, 18*(3), 317-348.

Kuhn, D. (1991). *The skills of argument.* New York: Cambridge University Press.

Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review, 62*(2), 155-178.

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education, 77*(3), 319-377.

Kuhn, D., Schauble, L., & Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning. *Cognition and Instruction, 9,* 285-327.

Laidlow, E. N., Skok, R. L., & McLaughlin, T. F. (1993). The effect of note-taking and self-questioning on quiz performance. *Science Education, 77,* 75-82.

Langer, J. A., & Applebee, A. (1987). *How writing shapes thinking: A study of teaching and learning.* Urbana, IL: National Council of Teachers of English.

Linn, M. C., & Hyde, J. S. (1989). Gender, mathematics, and science. *Educational Researcher, 18,* 17-19, 22-27.

Lortie, S. (1975). *Schoolteacher: A sociological study.* University of Chicago Press.

Means, L. M., & Voss, J. F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels. *Cognition and Instruction, 14*(2), 139-178.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: A source book of new methods* (2nd ed.). Newbury Park, CA: Sage.

Miller, R., & Osborne, J. F. (2000). *Beyond 2000: Science education for the future.* London: King's College London.

Newell, G. E. (1984). Learning from writing in two content areas: A case study/protocol analysis. *Research in the Teaching of English, 18,* 265-287.

Newton, P., Driver, R., & Osborne, J. (1999). The place of argument in the pedagogy of school science. *International Journal of Science Education, 21,* 553-576.

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching, 41*(10), 994-1020.

Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology, 49,* 31-57.
Scott, P. (1998). Teacher talk and meaning making in science classrooms: A Vygotskian analysis and review. *Studies in Science Education, 32,* 45-80.
Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking.* New York and Oxford: Oxford University Press.
Suppe, F. (1998). The structure of a scientific paper. *Philosophy of Science, 65,* 381-405.
Tierney, R. (1981). Using expressive writing to teach biology. In A. M. Wotring & R. Tierney (Eds.), *Two studies on writing in high school science (Classroom Research Study No. 5)* (pp. 47-79). Berkeley, CA: University of California, Bay Area Writing Project.
Toulmin, S. (1958). *The uses of argument.* Cambridge, England: Cambridge University Press.
Toulmin, S., Rieke, R., & Janik, A. (1984). *An introduction to reasoning* (2nd ed.). New York: Macmillan.
Zohar, A., & Dori, Y. J. (2003). Higher order thinking skills and low achieving students: Are they mutually exclusive? *Journal of the Learning Sciences, 12*(2), 145-182.
Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching, 39*(1), 35-62.

Correspondence regarding this article should be directed to:

Dr. Anita Roychoudhury
4138 Beering Hall
Curriculum and Instruction
Purdue University
West Lafayette, IN 47907-2098
Phone: (765) 496-3920
Fax: (765) 496-1622
aroychou@purdue.edu